APPROXIMATIONS BY ORTHOGONAL FUNCTIONS IN CASUALTY INSURANCE

Operations Research Center Research Report No. 73-24

Pantelis M. Pechlivanides

October 1973

U. S. Army Research Office - Durham

DA-31-124-ARO-D-331

Operations Research Center
University of California, Berkeley

ACKNOWLEDGEMENT

I would like to thank Professor William S. Jewell for his ideas and encouragement.

ABSTRACT

The premium calculation in insurance is of great importance. Knowing past claims $\xi_1, \ldots, \xi_n$ one wishes to forecast the next claim ξ_{n+1}, or in a more general fashion, forecast a function of ξ_{n+1}, $f(\xi_{n+1})$.

Since we are interested in a least squares forecast of $f(\xi_{n+1})$, the problem readily reduces to calculating $E[f(\xi_{n+1}) \mid \xi_1, \ldots, \xi_n]$. This can formally be done by use of Bayes rule. As Bayes rule is difficult to apply the theory of approximation with an orthonormal set of functions is exploited in this paper, to reveal the nature of our problem. The credibility premium and credible distribution follow as partial results after a generalization of Ericson's Theorem.

APPROXIMATIONS BY ORTHOGONAL FUNCTIONS IN CASUALTY INSURANCE

by

Pantelis M. Pechlivanides

INTRODUCTION

Let $\underline{\xi} = \xi_1, \ldots, \xi_n$ i.i.d. random variables with density function $\rho_{\xi|\theta}(x \mid \theta)$, i.e., $\xi_i \sim \rho_{\xi|\theta}(x \mid \theta)$.

Let their joint density function be $\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta)$. Since they are independent $\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta) = \prod_{i=1}^{n} \rho_{\xi_i|\theta}(x_i \mid \theta)$.

Also let θ (where θ can be a vector) be distributed as $\rho_\theta(\theta)$. We will call this density the prior.

Now suppose that we have observed $\xi_1, \ldots, \xi_n$ (which correspond to the individual experience in insurance terminology) and we want to find a function $\phi(\underline{\xi})$ to approximate in the least square sense a function of the next observation $f(\xi_{n+1})$. The above is summarized as:

Find $\phi(\underline{\xi})$ so that

$$I = E_{\xi_{n+1},\underline{\xi}}[f(\xi_{n+1}) - \phi(\underline{\xi})]^2 \text{ is minimized.}$$

Where $E_{\xi_{n+1},\underline{\xi}}$ is the expectation w.r.t. density $\rho_{\underline{\xi},\xi_{n+1}}(\underline{x}_1, \ldots, x_{n+1}) = \int_\theta \rho_{\xi_{n+1}|\theta}(x_{n+1} \mid \theta)\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta)\rho_\theta(\theta)d\theta$.

The integral sign $\int$ is used in this paper to mean either integration or summation whichever is appropriate. Now I can be rewritten as:

$$I = E_{\underline{\xi}}\left[E_{\xi_{n+1}|\underline{\xi}}f^2(\xi_{n+1}) - \left(E_{\xi_{n+1}|\underline{\xi}}f(\xi_{n+1})\right)^2\right] + E_{\underline{\xi}}\left[E_{\xi_{n+1}|\underline{\xi}}f(\xi_{n+1}) - \phi(\underline{\xi})\right]^2$$

$$= \underbrace{E_{\underline{\xi}} \operatorname{Var}_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1})}_{I_0} + \underbrace{E_{\underline{\xi}}\left[E_{\xi_{n+1}|\underline{\xi}}f(\xi_{n+1}) - \phi(\underline{\xi})\right]^2}_{I_1} .$$

We can see that I is minimized over all possible functions $\phi(\underline{\xi})$[†] when $\phi(\underline{\xi}) = E_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1})$. Then $I_1 = 0$ and $I = I_0$, the standard error.

We will denote $E_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1}) = \Phi(\underline{\xi})$.

$\Phi(\underline{\xi})$ can be, at least in theory, calculated exactly by use of Bayes rule:

$$\Phi(\underline{x}) = \frac{\int_{\xi} f(y) \int_{\theta} \rho_{\xi_{n+1}|\theta}(y \mid \theta)\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta)\rho_{\theta}(\theta)d\theta dy}{\int_{\theta} \rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta)\rho_{\theta}(\theta)d\theta} .$$

As it is usually difficult to employ this formula, we try to find an approximation to $\Phi(\underline{\xi})$.

[†]Here and later in this paper we use alternately $\Phi(\underline{\xi})$ or $\Phi(\underline{x})$; the correct notation of course is the latter, but there is no danger of confusion.

I. APPROXIMATIONS USING ORTHONORMAL BASIS

Let X be the set of values $\underline{\xi}$ can take. The set of Lebesque square integrable functions on X form a Hilbert space with measure $dk_n(\underline{x}) \triangleq \rho_{\underline{\xi}}(\underline{x})d\underline{x}$ (where $d\underline{x} \triangleq dx_1, \ldots, dx_n$) and inner product defined as

$$<f_1(\underline{\xi}),f_2(\underline{\xi})> \triangleq \int_X f_1(\underline{x})f_2(\underline{x})dk_n(\underline{x}) = \int_X f_1(\underline{x})f_2(\underline{x})\rho(\underline{x})d\underline{x} .$$

Every Hilbert space has a complete orthonormal set of functions which is called a Basis.

Let the sequence $\{u_j(\underline{x})\}_{j=0}^{\infty}$ form a basis for this space then by Theory we know that we can write, for any function $\Phi(\underline{x})$ belonging to this space:

$$\Phi(\underline{x}) = \sum_{j=0}^{\infty} c_j u_j(\underline{x}) \quad \text{where} \quad c_j = <\Phi(\underline{x}),u_j(\underline{x})> .$$

Also let $\phi(\underline{x}) = \sum_{j=0}^{N} c_j u_j(\underline{x})$ then the c_j's that minimize $E_{\underline{\xi}}[\Phi(\underline{\xi}) - \phi(\underline{\xi})]^2$ are given by $c_j = <\Phi(\underline{x}),u_j(\underline{x})>$, i.e., they are the same as above and do not depend on N .

Before we go on we want to denote $h(\theta) \triangleq E_{\xi_{n+1}|\theta}f(\xi_{n+1})$ and note that:

$$\Phi(\underline{\xi}) = E_{\xi_{n+1}|\underline{\xi}}f(\xi_{n+1}) = E_{\theta|\underline{\xi}}E_{\xi_{n+1}|\underline{\xi},\theta}f(\xi_{n+1}) = E_{\theta|\underline{\xi}}E_{\xi_{n+1}|\theta}f(\xi_{n+1}) = E_{\theta|\underline{\xi}}h(\theta)$$

by independence of $\xi_{n+1},\underline{\xi}$ given θ .

Proposition 1:

Under the above notation

$$c_j = E_\theta[h(\theta)\cdot E_{\underline{\xi}|\theta}u_j(\underline{\xi})] \quad \text{where} \quad h(\theta) \triangleq E_{\xi_{n+1}|\theta}f(\xi_{n+1}) .$$

Proof:

$$c_j = \langle\Phi(\underline{\xi}), u_j(\underline{\xi})\rangle = E_{\underline{\xi}}\left[E_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1})\cdot u_j(\underline{\xi})\right] =$$

$$E_\theta E_{\underline{\xi}|\theta}\left[E_{\xi_{n+1}|\underline{\xi},\theta} f(\xi_{n+1})\cdot u_j(\underline{\xi})\right] .$$

$E_{\underline{\xi}|\theta}$ applies to each of the terms in the product separately since $E_{\xi_{n+1}|\underline{\xi},\theta} f(\xi_{n+1}) = E_{\xi_{n+1}|\theta} f(\xi_{n+1})$ and $\xi_{n+1}, \underline{\xi}$ are independent given θ . Thus

$$c_j = E_\theta\left[E_{\xi_{n+1}|\theta} f(\xi_{n+1})\cdot E_{\underline{\xi}|\theta} u_j(\underline{\xi})\right]$$

(1)

$$= E_\theta[h(\theta)\cdot E_{\underline{\xi}|\theta} u_j(\underline{\xi})] .$$

Q.E.D.

Discussion:

This result is quite general since $f(\xi_{n+1})$ can be "any" function (when the appropriate expectations have meaning) of ξ_{n+1} .

Some examples are (a) $f(\xi_{n+1}) = \xi_{n+1}^k$ if we are interested in forecasting the next claim $(k = 1)$ or a moment of it $(k > 1)$. (b) $f(\xi_{n+1}) = I(\xi_{n+1} - y)$ where

$$I(\xi_{n+1} - y) \triangleq \begin{cases} 1 & \xi_{n+1} \le y \\ 0 & \text{otherwise} . \end{cases}$$

To expand on (b) we note that if $f(\xi_{n+1}) = I(\xi_{n+1} - y)$ then $E_{\xi_{n+1}|\underline{\xi}} I(\xi_{n+1} - y) = 1\cdot P[\xi_{n+1} \le y \mid \underline{\xi}] + 0\cdot P[\xi_{n+1} > y \mid \underline{\xi}] = P[\xi_{n+1} \le y \mid \underline{\xi}] = E_{\theta|\underline{\xi}} P[\xi_{n+1} \le y \mid \underline{\xi}, \theta] = E_{\theta|\underline{\xi}} P[\xi_{n+1} \le y \mid \theta] = E_{\theta|\underline{\xi}} h(\theta)$ where $h(\theta) = P[\xi_{n+1} \le y \mid \theta]$. So we are approximating the distribution of $\xi_{n+1} \mid \underline{\xi}$ at a given point y . (c) For the discrete case one can set $f(\xi_{n+1}) = \delta(\xi_{n+1} - y)$ where

$$\delta(\xi_{n+1} - y) = \begin{cases} 1 & \xi_{n+1} = y \\ 0 & \text{otherwise} \end{cases}$$

then similarly as in (b) we have $E_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1}) = P[\xi_{n+1} = y \mid \underline{\xi}]$ and $h(\theta) = P[\xi_{n+1} = y \mid \theta]$.

The choice of a basis is not an easy task and is more or less an art in itself. Here we try to limit and discuss the several possibilities.

Let a sufficient statistic for θ exist. We call it $T = T(\underline{\xi})$.

Proposition 2:

(2) $$E_{\xi_{n+1}|\underline{\xi}} f(\xi_{n+1}) = E_{\xi_{n+1}|T(\underline{\xi})} f(\xi_{n+1}) .$$

Proof:

First we note that if a sufficient statistic $T_n(\underline{\xi})$ for θ (T_n and θ can be vectors) exists we can write $\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta) = \rho_{T_n|\theta}(t_n \mid \theta)\cdot H_n(\underline{x})$ where $\rho_{T_n|\theta}(t_n \mid \theta)$ is a density function of $T_n(\underline{\xi})$ and $H_n(\underline{x})$ a function of $x_1, \ldots, x_n$ alone. Thus,

(3) $$\rho_{\xi_{n+1},\underline{\xi}|\theta}(x_{n+1},\underline{x} \mid \theta) = \rho_{\xi_{n+1}|\theta}(x_{n+1} \mid \theta)\cdot\rho_{\underline{\xi}|\theta}(\underline{x} \mid \theta)$$

$$= \rho_{T_1(\xi_{n+1})|\theta}(t_1(x_{n+1}) \mid \theta)H_1(x_{n+1})\rho_{T_n(\underline{\xi})|\theta}(t_n(\underline{x}) \mid \theta)H_n(\underline{x})$$

Now:

$$\rho_{\xi_{n+1}|\underline{\xi}}(x_{n+1} \mid \underline{x}) = \frac{E_\theta\left[\rho_{\xi_{n+1},\underline{\xi}}(x_{n+1},\underline{x} \mid \theta)\right]}{E_\theta[\rho_{\underline{\xi}}(\underline{x} \mid \theta)]}$$

and by (3)

$$(4.1) \qquad = H_1(x_{n+1}) \frac{E_\theta\left[\rho_{T_1|\theta}(t_1(x_{n+1} \mid \theta)\cdot\rho_{T_n|\theta}(t_n(\underline{x}) \mid \theta)\right]}{E_\theta\rho_{T_n|\theta}(t_n(\underline{x}) \mid \theta)}$$

and

$$(4.2) \qquad \rho_{\xi_{n+1}|T_n}(x_{n+1} \mid t_n(\underline{x})) = \frac{E_\theta\left[\rho_{\xi_{n+1}|\theta}(x_{n+1} \mid \theta)\cdot\rho_{T_n|\theta}(t_n(\underline{x}) \mid \theta)\right]}{E_\theta\rho_{T_n|\theta}(t_n(\underline{x}) \mid \theta)}$$

but $\rho_{\xi_{n+1}|\theta}(x_{n+1} \mid \theta) = \rho_{T_1|\theta}(t_1(x_{n+1}) \mid \theta)\cdot H_1(x_{n+1})$ and thus (4.1) equals (4.2)

Q.E.D.

Because of the above proposition, there exists $\tilde{\Phi}(\cdot)$ such that $\Phi(\underline{\xi}) = \tilde{\Phi}(t(\underline{\xi}))$. That is, $\Phi(\underline{\xi})$ is a function of the sufficient statistics for θ only.

This limits our search for a basis to a set of independent functions of t, i.e., $u_j(\underline{\xi}) = w_j(t(\underline{\xi}))$. It is our choice to write either:

$$E_{\underline{\xi}}[w_j(t(\xi))\cdot w_i(t(\underline{\xi}))] = 0 \quad \text{density} \quad \rho_{\underline{\xi}}(\underline{x})$$

or

$$E_T[w_j(t)\cdot w_i(t)] = 0 \quad \text{density} \quad \rho_T(t)$$

whichever is more helpful.

When T is not a vector, one choice of basis is to use polynomials of the sufficient statistics T. This is of course possible only if the square integrality condition $\left(\int_T t^k \rho_T(t)dt < \infty, \ k \text{ integer}\right)$ is satisfied. This is true, for example, when $T \in [0,\infty)$ and $\rho_T(t)$ is of exponential form.

If we want to use polynomials in T as a basis, we can use a three term reoccurrence relation to generate them. This can be found in books on orthogonal

expansions [4]:

$$w_0(T) = 1$$

$$w_1(T) = T - E_T T$$

$$w_{k+1}(T) = Tw_k(T) - \alpha_{k+1} w_k(T) - \beta_{k+1} w_{k-1}(T) \quad , \; k > 1$$

where

$$\alpha_{k+1} = \frac{\int t w_k^2(t)\rho_T(t)dt}{\int w_k^2(t)\rho_T(t)dt}$$

$$\beta_{k+1} = \frac{\int t w_k(t) w_{k-1}(t)\rho_T(t)dt}{\int w_{k-1}^2(t)\rho_T(t)dt} .$$

A similar three term reoccurrence can be shown to exist between the c_j's by direct substitution of the three term reoccurrence of basis into the formula of Proposition 2

$$c_j = \frac{E_\theta\{[E_{T|\theta}(T\cdot w_{j-1}(T))]\cdot h(\theta)\}}{||w_{j-1}(T)||} - \alpha_j c_{j-1} - \beta_j c_{j-2} \quad \text{where} \quad ||w_j(T)|| \triangleq \sqrt{E_T[w_j(T)]^2}$$

where α_j , β_j were found in calculating the basis.

We want to note that as the calculation of a polynomial basis requires only the knowledge of the distribution of the collective $\rho_{\underline{\xi}}(x)$, the calculation of c_j's require both the distribution of θ and $T \mid \theta$ (or equivalently $\underline{\xi} \mid \theta$) .

II. THE ERROR IN OUR APPROXIMATION

We said in the beginning that $I = I_0 + I_1$. The best we can get (using Bayes rule) is $I_1 = 0$. Now $I_1 = E_{\underline{\xi}}[\Phi(\underline{\xi}) - \phi(\underline{\xi})]^2$. But

$$\Phi(\underline{\xi}) = \sum_{j=0}^{\infty} c_j u_j(\underline{\xi})$$

and

$$\phi(\underline{\xi}) = \sum_{j=0}^{N} c_j u_j(\underline{\xi}) .$$

Thus,

$$I_1 = E_{\underline{\xi}}\left[\sum_{j=N+1}^{\infty} c_j u_j(\underline{\xi})\right]^2 .$$

Proposition 3:

(5)
$$I_1 = \sum_{j=N+1}^{\infty} c_j^2 .$$

Proof:

We know that $E_{\underline{\xi}}\Phi^2(\underline{\xi}) = \langle\Phi(\underline{\xi}),\Phi(\underline{\xi})\rangle = \sum_{j=0}^{\infty} c_j^2$. Also by Pythagorean Theorem

$$\langle\Phi(\underline{\xi}),\Phi(\underline{\xi})\rangle = \sum_{j=0}^{N} c_j^2 + E_{\underline{\xi}}\left[\Phi(\underline{\xi}) - \underbrace{\sum_{j=0}^{N} c_j u_j(\underline{\xi})}_{\phi(\underline{\xi})}\right]^2$$

so $E_{\underline{\xi}}(\Phi(\underline{\xi}) - \phi(\underline{\xi}))^2 = \sum_{j=N+1}^{\infty} c_j^2$.

Q.E.D.

We want to note here that since $E_{\underline{\xi}}\Phi^2(\underline{\xi}) = \sum_{j=0}^{\infty} c_j^2 =$ constant, to minimize $I_1 = \sum_{j=N+1}^{\infty} c_j^2$ is equivalent to maximize $\bar{I}_1 = \sum_{j=0}^{N} c_j^2$.

This can provide us with a way of comparison of two bases, i.e., it can resolve a question like which of the two bases $\{u_j\}$ or $\{v_j\}$, that we are given, is better for an N-term approximation. We will prefer the one that maximizes $\bar{I}_1$.

III. TWO TERM APPROXIMATION

Here we restrict ourselves to two-term approximations of the form $\phi(\underline{\xi}) = c_0 u_0(\underline{\xi}) + c_1 u_1(\underline{\xi})$.

We pick $u_0(\underline{\xi}) = 1$ and

$$u_1(\underline{\xi}) = \frac{g(\underline{\xi}) - E_{\underline{\xi}} g(\underline{\xi})}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}}$$

where $g(\underline{\xi})$ any function of $\underline{\xi}$ that belongs to our space $\left(\text{i.e., } \int g(\underline{x})^2 \rho_{\underline{\xi}}(\underline{x}) d\underline{x} \equiv E_{\underline{\xi}} g^2(\underline{x}) < \infty\right)$.

First, we note that u_0 , u_1 are orthogonal:

$$<u_0, u_1> \equiv \frac{E_{\underline{\xi}}[g(\underline{\xi}) - E_{\underline{\xi}} g(\underline{\xi})]}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} = 0 \ .$$

u_1 is also normalized:

$$||u_1(\underline{\xi})|| = \frac{\sqrt{E_{\underline{\xi}}(g(\underline{\xi}) - E_{\underline{\xi}} g(\underline{\xi}))^2}}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} = 1 \ .$$

Now, using Proposition 1,

$$c_0 = E_\theta h(\theta)$$

$$c_1 = E_\theta[h(\theta) \cdot E_{\underline{\xi}|\theta} u_1(\underline{\xi})] = \frac{E_\theta[h(\theta) \cdot (E_{\underline{\xi}|\theta} g(\underline{\xi}) - E_{\underline{\xi}} g(\underline{\xi}))]}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} =$$

$$= \frac{E_\theta h(\theta) \cdot E_{\underline{\xi}|\theta} g(\underline{\xi}) - E_\theta h(\theta) \cdot E_\theta E_{\underline{\xi}|\theta} g(\underline{\xi})}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} = \frac{\mathrm{Cov}_\theta\ [h(\theta); E_{\underline{\xi}|\theta} g(\underline{\xi})]}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} \ .$$

Thus,

$$(6) \qquad \phi(\underline{\xi}) = E_\theta h(\theta) + \frac{\mathrm{Cov}_\theta\ [h(\theta);E_{\underline{\xi}|\theta}g(\underline{\xi})]}{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}\ [g(\underline{\xi}) - E_{\underline{\xi}}g(\underline{\xi})] \ .$$

We call

$$a = \frac{\mathrm{Cov}_\theta\ [h(\theta);E_{\underline{\xi}|\theta}g(\underline{\xi})]}{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})} \ .$$

Thus,

$$a = \frac{c_1}{\sqrt{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})}} \ .$$

From the preceding analysis, a generalization of Ericson's Theorem follows: (we keep the same notation)

Theorem:

Let $\underline{\xi} = \xi_1, \ldots, \xi_n$ be independent identically distributed random variables conditional on θ . Let $h(\theta)$ be a function of θ and $g(\underline{\xi})$ an unbiased estimate of $h(\theta)$, i.e., $E_{\underline{\xi}|\theta}g(\underline{\xi}) = h(\theta)$.

If we want to approximate $E_{\theta|\underline{\xi}}h(\theta)$ by $b + ag(\underline{\xi})$, the least square coefficients a , b are given by:

$$(7) \qquad a = \frac{\mathrm{Var}_\theta\ h(\theta)}{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})} \ , \ b = (1 - a)E_\theta h(\theta) \ .$$

Proof:

Follows directly by (6) by setting $E_{\underline{\xi}|\theta}g(\underline{\xi}) = h(\theta)$

Q.E.D.

Now we note that

(8) $$\mathrm{Var}_{\underline{\xi}}\, g(\underline{\xi}) = E_T\, \mathrm{Var}_{\underline{\xi}|T}\, g(\underline{\xi}) + \mathrm{Var}_T\, E_{\underline{\xi}|T} g(\underline{\xi})$$

and that if $T = T(\underline{\xi})$ is the sufficient statistic of θ for $\rho(\underline{x} \mid \theta)$ then $L(t(\underline{\xi})) \triangleq E_{\underline{\xi}|T} g(\underline{\xi})$ is still an unbiased estimator of $h(\theta)$ because $E_{\underline{\xi}|\theta} E_{\underline{\xi}|T} g(\underline{\xi}) = E_{T|\theta} E_{\underline{\xi}|T,\theta} g(\underline{\xi})$ since $E_{\underline{\xi}|T,\theta} g(\underline{\xi})$ is a function of T only, if T is a sufficient statistic for θ. Thus, $E_{\underline{\xi}|\theta} E_{\underline{\xi}|T} g(\underline{\xi}) = E_{\underline{\xi}|\theta} g(\underline{\xi}) = h(\theta)$. Also $\mathrm{Var}_{\underline{\xi}}\, L(t(\underline{\xi})) \leq \mathrm{Var}_{\underline{\xi}}\, g(\underline{\xi})$ by (8), i.e., $L(t(\underline{\xi}))$ has minimum variance over all unbiased estimators. This is an improvement over $g(\underline{\xi})$ since

$$c_1^2 = \frac{(\mathrm{Var}_\theta\, h(\theta))^2}{\mathrm{Var}_{\underline{\xi}}\, g(\underline{\xi})}$$

is maximized when $g(\underline{\xi}) = L(t(\underline{\xi}))$. Thus, $\bar{I}_1 = c_0^2 + c_1^2$ is maximized. So in theory at least, we can find the best two-term expansion within the class of unbiased estimators of $h(\theta)$.

The question is how to find $L(t(\underline{\xi}))$:

1. We find an unbiased estimator for $h(\theta)$. This is readily available. Take $f(\xi_1)$. Remember we defined $h(\theta) = E_{\xi|\theta} f(\xi)$. Thus, $f(\xi_1)$ is an unbiased estimator for $h(\theta)$.
2. Find a sufficient statistic $T(\underline{\xi})$ for θ and the density function of T, $\rho_{T|\theta}(t \mid \theta)$.
3. Calculate $E_{\underline{\xi}|T} f(\xi_1)$ using $\rho_{T|\theta}(t \mid \theta)$, $\rho_\theta(\theta)$ if available.

It is obvious that Steps 2 and 3 can be very difficult in practice. Finally, we remark that for the exponential family the sufficient statistic is complete; thus, the min variance unbiased estimator is unique.

<u>Example I</u>:

We want to calculate the net premium. Then $f(\xi_{n+1}) = \xi_{n+1}$. We will use Ericson's Theorem. Let us choose $g(\underline{\xi}) = \bar{\xi} \equiv \frac{\sum_{i=1}^{n} \xi_i}{n}$ which is unbiased, i.e.,

$E_{\underline{\xi}|\theta} g(\underline{\xi}) = E_{\underline{\xi}|\theta} \bar{\xi} = E_{\xi|\theta} \xi = h(\theta)$. Then

$$a = \frac{\mathrm{Var}_\theta\ h(\theta)}{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})} = \frac{E_\theta[(h(\theta))^2] - (E_\xi \xi)^2}{\frac{1}{n} E_\theta\ \mathrm{Var}_{\xi|\theta} \xi + \mathrm{Var}_\theta\ h(\theta)}$$

which is exactly the formula found in Bühlman [1].

Example II:

For the variance principle in premium calculation, we need to approximate the quantity $\mathrm{Var}_{\xi_{n+1}|\underline{\xi}} \xi_{n+1}$

$$\mathrm{Var}_{\xi_{n+1}|\underline{\xi}} \xi_{n+1} = E_{\xi_{n+1}|\underline{\xi}}\left(\xi^2_{n+1}\right) - \left(E_{\xi_{n+1}|\underline{\xi}} \xi_{n+1}\right)^2 .$$

Noting that ξ^2_{n+1} is a function of ξ_{n+1} , $f(\xi_{n+1})$, we can apply previous theory on orthonormal basis to approximate $E_{\xi_{n+1}|\underline{\xi}} \xi^2_{n+1}$ or use a two-term expansion.

As for the second term, an approximation can be found by squaring the approximation of the net premium.

Example III:

We want to approximate $P[\xi_{n+1} \le y \mid \underline{\xi}]$. Then $f(\xi_{n+1}) = I(\xi_{n+1} - y)$ and $h(\theta) = P[\xi_{n+1} \le y \mid \theta] = P[y \mid \theta]$. Applying Ericson's Theorem, we choose

$$g(\underline{\xi}) = \frac{1}{n} \sum_{i=1}^{n} I(\xi_i - y)$$

which is unbiased, i.e., $E_{\underline{\xi}|\theta} g(\underline{\xi}) = h(\theta)$. Then

$$a = \frac{\mathrm{Var}_\theta\ h(\theta)}{\mathrm{Var}_{\underline{\xi}}\ g(\underline{\xi})} = \frac{\mathrm{Var}_\theta\ [P(y \mid \theta)]}{E_\theta\ \mathrm{Var}_{\underline{\xi}|\theta} \frac{1}{n} \sum I(\xi_i - y) + \mathrm{Var}_\theta\ E_{\underline{\xi}|\theta} \frac{1}{n} \sum I(\xi_i - y)} .$$

The denominator can be written as

$$E_\theta\left[\frac{1}{n} P(y \mid \theta)(1 - P(y \mid \theta))\right] + \mathrm{Var}_\theta\, P(y \mid \theta) =$$

$$= E_\theta\left[\frac{1}{n} P(y \mid \theta)\right] - \frac{1}{n} E_\theta P^2[y \mid \theta] + \mathrm{Var}_\theta\, P(y \mid \theta) =$$

$$= \frac{1}{n} P(y) - \frac{1}{n} \mathrm{Var}_\theta\, P(y \mid \theta) - \frac{1}{n} (E_\theta P(y \mid \theta))^2 + \mathrm{Var}_\theta\, P(y \mid \theta) =$$

$$= \frac{1}{n} P(y)(1 - P(y)) + \left(1 - \frac{1}{n}\right) \mathrm{Var}_\theta\, P(y \mid \theta)$$

where

$$P(y) = E_\theta P(y \mid \theta) .$$

This is exactly the formula for the credible distribution found by Jewell [7].

IV. A GENERALIZATION TO MANY DIMENSIONS

We can easily replace ξ_{n+1} by a vector $\vec{\xi}_{n+1}$ and $\underline{\xi}$ by a matrix Ξ and $f(\xi_{n+1})$ by $\vec{f}(\vec{\xi}_{n+1})$

$$\vec{f}(\vec{\xi}_{n+1}) = \begin{bmatrix} f_1(\xi_{n+1}) \\ \cdot \\ \cdot \\ \cdot \\ f_m(\xi_{n+1}) \end{bmatrix} .$$

The relevant density is now $\rho_{\Xi|\theta}(X \mid \theta) = \prod_{t=1}^{n} \rho_{\vec{\xi}_t|\theta}(\vec{x}_t \mid \theta)$

$$I = E_{\xi_{n+1},\Xi}[\vec{f}(\vec{\xi}_{n+1}) - \vec{\phi}(\Xi)]^2 = E_{\xi_{n+1},\Xi}\left[\sum_{i=1}^{m} [f_i(\vec{\xi}_{n+1}) - \phi_i(\Xi)]^2\right] .$$

So the problem of finding $\phi_i(\Xi)$, $i = 1, \ldots, m$ to minimize I can be split to m subproblems: Find $\phi_i(\Xi)$ to minimize I_i

$$I_i = E_{\xi_{n+1},\Xi}(f_i(\vec{\xi}_{n+1}) - \phi_i(\vec{\Xi}))^2 \ , \ i = 1, \ldots, m .$$

Now I_i can be written as $I_i = I_{i0} + I_{i1}$ where we define

$$I_{i0} = E_{\Xi}\left[E_{\vec{\xi}_{n+1}|\Xi} f_i^2(\vec{\xi}_{n+1}) - \Phi_i^2(\Xi)\right]$$

$$I_{i1} = E_{\Xi}[\Phi_i(\Xi) - \phi_i(\Xi)]^2$$

and

$$\Phi_i(\Xi) \triangleq E_{\vec{\xi}_{n+1}|\Xi} f_i(\vec{\xi}_{n+1}) .$$

Similar results should follow as before with the appropriate complication.

Example:

Suppose $\vec{f}(\vec{\xi}_{n+1}) = f(\vec{\xi}_{n+1}) = I(\vec{\xi}_{n+1} - \vec{y})$ where

$$I(\vec{\xi}_{n+1} - \vec{y}) = \begin{cases} 1 & \vec{\xi}_{n+1} \leq \vec{y} \\ 0 & \text{otherwise.} \end{cases}$$

We choose to make a two-term approximation to $E_{\vec{\xi}_{n+1}|\Xi} f(\vec{\xi}_{n+1})$.

We wish to use Ericson's Theorem so we restrict ourselves to unbiased estimator of $h(\theta) = E_{\vec{\xi}_{n+1}|\theta} f(\vec{\xi}_{n+1}) = P[\vec{\xi}_{n+1} \leq \vec{y} \mid \theta]$. One such is readily available. Take

$$g(\Xi) = \frac{1}{n} \sum_{i=1}^{n} I(\vec{\xi}_i - \vec{y}_i) .$$

Then

$$a = \frac{\mathrm{Var}_\theta\, h(\theta)}{\mathrm{Var}_\Xi\, g(\Xi)} = \frac{\mathrm{Var}_\theta\, P[\vec{\xi}_{n+1} \leq \vec{y} \mid \theta]}{E_\theta\, \mathrm{Var}_{\Xi|\theta} \dfrac{\sum I(\vec{\xi}_2 - \vec{y}_i)}{n} + \mathrm{Var}_\theta\, E_{\Xi|\theta} \dfrac{\sum_{i=1}^{n} I(\xi_i - y_i)}{n}} =$$

by similar calculation as in Example III

$$= \frac{\mathrm{Var}_\theta\, P[\vec{y} \mid \theta]}{\frac{1}{n} P(\vec{y})(1 - P(\vec{y})) + \left(1 - \frac{1}{n}\right) \mathrm{Var}_\theta\, P(\vec{y} \mid \theta)}$$

where $P(\vec{y}) = P[\vec{\xi}_t \leq \vec{y}]$ and $P[\vec{y} \mid \theta] = P[\vec{\xi}_t \leq \vec{y} \mid \theta]$.

V. AN ACADEMIC APPLICATION

This paper's purpose is not to advocate the use of orthogonal functions to approximate the Bayes rule as it appears on Page 2.

We used this approach to unite the idea of credibility in insurance whether this is for the approximation of the mean or the credible distribution or any other function as it was explained before.

In view of the calculations involved in approximating the Bayes rule by orthonormal functions, the question still exists--what are the advantages, if any, from the computational point of view, to calculating the Bayes formula directly.

We can say the following:

(a) Directly from the Bayes formula, we have to do a number of integrations to find $\Phi(\underline{x})$ for a specific value of $\underline{x}$. An improvement to that is to find $\tilde{\Phi}(t_n)$ (a function of the sufficient statistics T_n)

$$\tilde{\Phi}(t_n) \equiv E_{\theta|T_n} E_{\xi|\theta} f(\xi) .$$

Again, our effort will result in finding (with the use of a computer) the value of $\tilde{\Phi}(t_n)$ at a specific value of t_n .

(b) Using orthogonal functions of the sufficient statistics for θ , $\left\{w_j(t_n)\right\}_{j=0}^{\infty}$, we make initially a bigger effort to find them numerically. After that, given any function $f(\xi_{n+1})$ for which $h(\theta) \equiv E_{\xi|\theta} f(\xi)$, we use the same $\{w_j(t_n)\}$ to find the coefficients $\{c_j\}$ according to formula (1) for each $h(\theta)$, our goal being to form the sum $\sum_{j=0}^{N} c_j w_j(t_n)$ which is our approximation to the Bayes formula. But this sum is a function of t_n . So no extra integrations are needed for another customer with different t_n . (Note n must remain the same.)

The Example

We choose

$$\rho_{\xi|\theta}(x \mid \theta) = \frac{1}{\sqrt{2\pi}\, v} \exp\left(-\frac{(x-\theta)^2}{2v^2}\right) = N(\theta, v^2)$$

$$\rho_\theta(\theta) = N(\mu, \sigma^2) .$$

A sufficient statistic for θ is $T_n = \sum_{i=1}^{n} \xi_i$ and it is easily shown that

$$\rho_{T_n|\theta}(t_n \mid \theta) = N(n\theta, nv^2) .$$

It follows that

$$\rho_{T_n}(t_n) = N(n\mu, n(n\sigma^2 + v^2)) .$$

Numerically, the first seven orthonormal polynomials in t w.r.t. density $\rho_{T_n}(t_n)$ were found with $n = 10$, $\mu = 2$, $\sigma^2 = 2$, $v^2 = 5$:

j	t^0	t^1	t^2	t^3	t^4	t^5	t^6
0	1.						
1	-1.27	.064					
2	.473	-.12	.003				
3	.746	.071	.008	$.12 \times 10^{-3}$			
4	-1.04	.085	.007	$.43 \times 10^{-3}$	$.53 \times 10^{-5}$		
5	.170	-.191	$.57 \times 10^{-2}$	$.52 \times 10^{-3}$	$.23 \times 10^{-4}$	$.22 \times 10^{-6}$	
6	.854	.082	-.019	$.22 \times 10^{-3}$	$.350 \times 10^{-4}$	$.11 \times 10^{-5}$	$.91 \times 10^{-8}$

TABLE 1: THE ORTHONORMAL POLYNOMIALS $\left\{w_j(t_{10})\right\}_{j=0}^{6}$

Next we calculate the coefficients $\left\{c_j\right\}_{j=0}^{6}$ for three different functions:

(a) $h_1(\theta) \equiv E_{\xi|\theta}\xi$

(b) $h_2(\theta) \equiv E_{\xi|\theta}\xi^2$

(c) $h_3(\theta) \equiv P[\xi < x \mid \theta]$, $x = 2$.

They are as follows:

	c_0	c_1	c_2	c_3	c_4	c_5	c_6	$\sum_{j=0}^{6} c_j^2$
$h_1(\theta)$	1.99	1.05	-.156	-.016	-.115	-.073	.099	5.135
$h_2(\theta)$	10.892	3.96	1.43	-.571	-.218	-.162	1.596	139.293
$h_3(\theta)$	.490	-.232	.029	.014	.011	-.051	.065	.3026

TABLE 2

Looking at the sum of squares of the c_j's , we see that in all cases the main contribution comes from the first two terms. This can be used also as an indication that a two term approximation with $g(\underline{x}) = t(x)$ would be appropriate (see Section III). Specifically,

(a) For $h_1(\theta) = \theta$ we know that the correct result is

$$c_0 + c_1\left(\frac{\frac{t_n}{n} - c_0}{\sqrt{\operatorname{Var} \frac{T_n}{n}}}\right) \qquad t_n = \sum_{i=1}^{n} x_i \text{ , } n = 10$$

$$c_0 = E_\theta h(\theta) = 2$$

$$c_1 = \frac{\mathrm{Var}_\theta\ h(\theta)}{\sqrt{\mathrm{Var}\ \frac{\sum x_i}{n}}} = \frac{\sigma^2}{\sqrt{\frac{1}{n^2}\, n(n\sigma^2 + v^2)}} = \frac{2 \times \sqrt{10}}{\sqrt{20+5}} \approx 1.24$$

$$c_j = 0\ ,\ j > 1\ .$$

These results are very close to our estimations of c_0 , c_1 in Table 2. The rest of the c_j's are one order of magnitude smaller, thus their contribution to the sum of squares is minimal.

(b) For $h_2(\theta) = v^2 + \theta^2$ the correct result is

$$E_{\xi_{n+1}|T_n} \xi_{n+1}^2 = v^2 + \frac{\sigma^2 v^2}{v^2 + n\sigma^2} + \left(\frac{\mu v^2 + \sigma^2 t}{v^2 + n\sigma^2}\right)^2$$

setting $v^2 = 5$, $\sigma^2 = 2$, $\mu = 2$

$$= 5.5 + \frac{40}{625}\, t + \frac{4}{625}\, t^2$$

which is also very close to what we find by using orthonormal polynomials

(c) Finally, for $h_3(\theta) = P[\xi \le x \mid \theta]$, $x = 2$, one can show that $\rho_{\xi_{n+1}|T}(x \mid t) = N(\mu', u^2 + v^2)$ where

$$\mu' = \frac{\mu v^2 + \sigma^2 t}{v^2 + n\sigma^2} = .4 + \frac{2}{25}\, t$$

and

$$u^2 = \frac{\sigma^2 v^2}{v^2 + n\sigma^2} = \frac{27}{5}\ .$$

Using tables for Normal distribution, we find what appears in Table 3 under column "correct value".

Now using orthonormal polynomials in t , we have:

Two-Term Approximation:

$$P[\xi \leq 2 \mid T = t] \quad .78464 - .01484t \ .$$

Three-Term Approximation:

$$P[\xi \leq 2 \mid T = t] \quad .798357 - .018328t + .000087t^2 \ .$$

The calculations are summarized in the following table:

t_{10}	Two-Term Approximation	Three-Term Approximation	Correct Value
10	.636	.624	.637
20	.488	.467	.500
30	.339	.327	.363
40	.191	.204	.242
50	.04224	.099	.1482

TABLE 3: $P[\xi \leq 2 \mid T_{10} = t_{10}]$

We see that for big values of t we need more terms to have an acceptable approximation.

Finally, we want to remark that the computer effort involved (not including programming effort) was minimal in this particular example since we chose $\text{Var}\ \theta = \sigma^2 = 2$. If σ^2 were big, a larger range of integration over θ would b needed thus increasing the computer effort vastly since we are dealing with double integrals over x and θ .

REFERENCES

[1] Bühlman, H., MATHEMATICAL METHODS IN RISK THEORY, Springer-Verlag, New York, (1970).

[2] DeGroot, M. H., OPTIMAL STATISTICAL DECISIONS, McGraw-Hill, New York, (1970).

[3] Ferguson, T. S., MATHEMATICAL STATISTICS: A DECISION THEORETIC APPROACH, Academic Press, New York, (1967).

[4] Hamming, R. W., NUMERICAL METHODS FOR SCIENTISTS AND ENGINEERS, McGraw-Hill, (1962).

[5] Hogg, R. V. and A. T. Craig, INTRODUCTION TO MATHEMATICAL STATISTICS, Macmillan, 3rd Edition, (1970).

[6] Jewell, W. S., "Multi-Dimensional Credibility," ORC 73-7, Operations Research Center, University of California, Berkeley, (April 1973).

[7] Jewell, W. S., "The Credible Distribution," ORC 73-13, Operations Research Center, University of California, Berkeley, (August 1973).

[8] Reed, M. and B. Simon, FUNCTIONAL ANALYSIS, Vol. 1, Academic Press, New York, (1972).

[9] Salmond, D., "Premium Calculation in Casualty Insurance," ORC 73-2, Operations Research Center, University of California, Berkeley, (February 1973).